craic computing

Select Agent Program
Centers for Disease Control and Prevention
1600 Clifton Road NE
Mail Stop E79
Atlanta, GA 30333

18th February, 2003

Dear Colleagues,

I wish to comment on 42 CFR Part 73, Interim Final Rule, Possession, Use, and Transfer of Select Agents and Toxins published in the Federal Register, Vol. 240, No. 67 on Friday, December 13, 2002.

The new regulations represent a significant improvement over 42 CFR Part 72 with regard to genetic material derived from select agents. Sections 73.4(e) and 73.5(e) provide specific scenarios where genetic materials from select agents are subject to the regulations. However, there are many other ways in which sequences from these agents can be manipulated and which could have serious implications for biological safety.

An extraordinary range of technologies are available for genetic manipulation. Any attempt to encompass that diversity in the limited terms of these regulations is virtually impossible. This creates a gray area in which work involving sequences from these agents can proceed without regulation, with the potential for consequences that the regulations seek to control. For example, the synthesis of short oligonucleotides for sequences from any agent is not covered by the regulations. Whereas the synthesis of the entire genome of a select agent or the gene encoding a specific toxin, using these oligonucleotides, is covered.

Faced with a set of regulations one can interpret them rigidly, as evidenced by the comments of the DOE Joint Genome Institute and Lawrence Berkeley National Laboratory in their submission ref: 03-SAR-043. Here they interpret the specific terms of the new regulations such that they do not need to register their work involving sequencing of bacterial select agents. A liberal reading of the intent behind the regulations would suggest that this work should be subject to control. It this ambiguity that is of concern to us.

The ambiguity poses a serious problem for biotechnology service companies that provide oligonucleotide and gene synthesis and DNA sequencing services to other companies as well as academic and government groups. In particular, these companies typically are not informed by their clients of the precise nature of the sequences. Unwittingly their work may be used in a later activity that is covered by the regulations. A striking example of how this could occur is the use of commercial produced oligonucleotides in the synthesis of the poliovirus genome and subsequent production of infective virus particles (Wimmer et al., Science 2002, v297, pp1016-1018). There is concern among a number of these companies that the current regulations do not address their situation. Without regulation or guidance from the CDC and other agencies, these companies face serious issues of liability if their services are misused in the manipulation of select agents by others. Additional background material on this issue is given in the 'white paper' that I have included below.

The CDC, USDA, Dept of Commerce and other relevant agencies should enter into a dialog with this sector of the biotechnology industry to explore this gray area in the control of hazardous biological agents.

Sincerely,

Robert Jones, D.Phil, President, Craic Computing LLC

Craic Computing LLC, 911 East Pike St. #208, Seattle, WA 98122

Phone: 206-860-7118 Fax: 206-860-7119 Email: jones@craic.com

Hazardous Biological Agents, Biotechnology Companies and DNA Sequences - A Background Paper

The increased threat of biological terrorism in the US following the anthrax attacks of 2001 has forced the biotechnology industry to review how it handles pathogens and other hazardous biological agents. In this background paper we describe the issues that are faced by one sector of that industry and show how they pose a unique set of problems that demand attention.

Our company, Craic Computing LLC, provides custom software and data analysis services to biotechnology companies. A number of our clients face the issues described here and in response we have created a software package that can provide a partial response to these needs. We are eager to promote a dialog between the industry and the government agencies to better explore the potential threats and to develop novel approaches to dealing with them. We hope this paper will contribute to this debate.

Biotechnology Service Companies

Within the biotechnology industry a significant number of companies provide contract services involving the manipulation of DNA to clients in industrial, academic and federal organizations. Three of the major services are oligonucleotide synthesis, gene synthesis and sequencing.

- **Oligonucleotide synthesis** companies chemically synthesize short pieces of DNA for a client based on the DNA sequences they provide. This is typically highly automated with facilities synthesizing thousands of molecules every day. This represents the largest of the three services and it is highly competitive.
- **Gene synthesis** is a related technology that constructs entire genes out of shorter oligonucleotides, based on a supplied DNA sequence. The chemical synthesis is followed by gene insertion in a plasmid vector and transformation into bacteria and, typically, confirmatory sequencing of the synthetic gene. This is a smaller component of the market.
- **DNA sequencing** companies determine the sequence of DNA supplied to them as native DNA or as genetic constructs in bacteria or viruses. Whereas the synthesis companies start with a DNA sequence and produce genetic material, sequencing companies generate sequence data from DNA. These companies comprise a large portion of the market.

The Problem

There is growth in this sector of the industry as new genomic technologies are applied ever more widely. Companies compete fiercely in terms of cost, ease of access and turnaround time for their services. In the area of oligonucleotide synthesis the major vendors all provide web-based ordering, credit card payment and shipment of the DNA within 24 hours. This unprecedented access to custom reagents has revolutionized molecular biology. But the convenience to the customer can create hidden risks for the supplier.

In order to protect their intellectual property, biotechnology companies are very reluctant to disclose any details of their research to a third party. Service companies realize this and rarely, if ever, ask about the nature or intended uses of the sequence data or DNA samples that are supplied to them. This intentional lack of disclosure is critical to the way service companies operate. Without it many of their customers would choose to meet these needs in-house.

But this ignorance about the materials they work with can have serious implications for the service company. Inadvertently, their staff may be exposed to hazardous biological agents for which they are totally unprepared. They might contravene federal regulations regarding transport and possession of hazardous agents without realizing it. And, albeit unlikely, a client might use their services for an illicit purpose, the most dramatic of which would involve biological terrorism.

Most researchers believe that the intentional modification of a hazardous biological agent for purposes of terrorism is an unlikely scenario. There are many ways to misuse these naturally occurring agents without resorting to the complexity of genetic engineering. But in the current atmosphere an argument like that is no longer adequate. The engineering of hazardous biological agents has occurred intentionally in Soviet bioweapons programs and unintentionally in academic labs. The public is aware of the scenario and is concerned about it. For that reason alone, service companies need to address the issue of hazardous biological agents and how they affect their business.

Examples of the Problem

The following real world examples show how legitimate research can expose service companies and their staff to potential risk. These alone should be sufficient cause for concern.

Polio

In August 20002, Dr. Wimmer et al. from SUNY reported the synthesis of the complete genome of Poliovirus and the successful production of infectious virus from that DNA (Science 2002, v297, pp 1016-1018). The genome was assembled from oligonucleotides made by Integrated DNA Technologies of Iowa. This report attracted significant media attention and sparked controversy in the scientific community and in government.

This demonstrates how a service company can unwittingly be drawn into work on a hazardous agent.

Mousepox

As part of a legitimate research project at CSIRO in Australia, Dr. Jackson et al. (J. Virology 2001, v75, pp1205-1210) inserted a gene for Interleukin-4 into mousepox virus. This had the unexpected effect of suppressing the normal immunological response from the host, resulting in significantly increased virulence. The engineered virus was even able to overcome the protection offered by vaccination. This report raised considerable concern within the virology community.

This demonstrates how genetic manipulations involving pathogens can have unexpected results and provides a specific example of a manipulation that increases virulence.

Blue Heron Biotechnology

The experiences of Blue Heron Biotechnology, a gene synthesis company in Bothell, WA, illustrate how the issues described here impact the operation of a service company. Craic Computing is working closely with Blue Heron to help screen for sequences from hazardous agents submitted to them for synthesis. This approach has identified two cases that raised concerns for the company. In the first, a client requested that a portion of cholera toxin be optimized for expression in an edible plant. The company was concerned that the production of an edible plant expressing toxin genes could have unintended consequences. In the second, a client outside the US submitted the sequence for a gene from a Variola virus, related to smallpox. This legitimate request was not thought to have biological safety concerns but would have required the company to obtain an export license for the construct in order to comply with federal laws. Neither of these two issues would have been discovered were it not for the company taking the initiative and screening the sequences supplied to it.

These demonstrate that significant potential risks are being encountered today in biotechnology service companies.

The Four Key Issues

From this underlying concern with hazardous agents we can extract four keys issues that service companies need to address. We present them here, based on our overall assessment and opinion, in decreasing order of importance. Others may well view the relative importance differently.

1. Biological Safety

Working with sequences and materials of unknown origin can expose company staff to immediate health risks. We believe this risk is real, it is present and it needs to be addressed immediately. A company needs a way to identify potential risks either on submission of sequence data from the client or as soon as that data is generated in-house.

Failure to ensure a safe working environment has major personal, business and legal implications.

2. Legal Compliance

A company can not ensure that it is complying with federal laws without understanding the nature of the materials they work with. The relevant laws relate to possession, handling and transport of materials within the US and their export from, and import into, the country.

Failure to comply with these laws can result in severe financial and other penalties being imposed on a company and its officers.

3. Due Diligence

A company has a responsibility to its staff, its board and its shareholders to address each of the other three issues. This includes taking whatever practical measures are available to identify risks, to protect against them and to inform the relevant authorities in case of an incident. By being seen to be pro-active in these issues, a company not only demonstrates its diligence but may also deter its services being exploited for illicit purposes.

Failure to show due diligence, especially in the face of an incident, can mean financial ruin for a business.

4. Biological Terrorism

We view the actual exploitation of gene synthesis, etc. for purposes of terrorism as very unlikely. However, the implications of such an event are immense. Therefore the issue has to be addressed at some level appropriate to the threat. In addition, given the rapid evolution of technology and the recent history of terrorism, we need to prepare for the future when such an act is less unlikely.

Failure to address this threat has the low probability of immense harm to the public at large.

Solutions

There are several approaches that can address these issues. None of them is a complete solution.

Full disclosure by the Client

Much of the problem for service companies stems from the confidential nature of their relationship with their clients. If clients were required to disclose the nature of their sequences and their intended use then service companies could make an informed decision on whether to work on each project. However the impact of such disclosure on the service business would be immense. Biotechnology companies have a responsibility to protect their intellectual property and disclosing that information to a third party in any form would require a properly executed confidentiality agreement between the company and the service company. The cost and effort required would make the system unworkable and would significantly delay research. The result would be that the larger clients would perform these services in-house at considerably greater expense.

An alternative would involve partial disclosure in the form of an agreement that the client would not submit any sequences or materials that involve hazardous agents. Our colleagues at Blue Heron Biotechnology require customers to include notification with any order for sequences from select agents. But these only work if they are adhered to by both parties. Negligence or deceit on the part of the client, or ambiguity in the interpretation of the agreement, can render it useless. They are worthwhile but they are insufficient.

Legislation

Regulations such as those governing the possession and transfer of select agents can play an important role in biodefense and biological safety. These can provide clear guidance where the issue can be clearly delineated, such as the possession of smallpox virus. But they have difficulty encompassing the diversity of sequences and

reagents that might be derived from a biological agent or that might be used in its modification. The updated regulations on select agents (42 CFR Part 73) issued in December 2002 have begun to address DNA sequences derived from these agents but significant ambiguity remains. It is unlikely that legislation can provide complete control over the issues without seriously impacting legitimate research.

Detection

One way for a service company to protect itself from hazardous agents would be to screen for their presence in the reagents supplied to them or generated by them. But the diversity in the agents precludes any realistic attempt at this. However we can provide some level of detection if we examine DNA sequences. The three types of service company described here all use DNA sequence information, either supplied by the client for synthesis or generated by the company for the client. By comparing these data against a database of sequences from the most hazardous pathogens and toxins, we can detect the presence of these agents and assess the risk they pose. This strategy allows an oligonucleotide or gene synthesis to assess all sequences submitted to them prior to any laboratory work, thereby avoiding any health implications. For companies involved in DNA sequencing the screening would take place after staff have worked with the biological reagents. This is not ideal but it does allow any issues to be uncovered as soon as possible after the work has started.

The use of a custom database of pathogen sequences has two advantages over a broader database. It improves search speed by virtue of its relatively small size, but more importantly it protects the confidentiality of the client data by not uncovering any match to a database sequence other than that of a pathogen. This would seem to be an important compromise that can greatly help in the adoption of this approach. The service company is able screen for all the sequences that pose a risk to them and yet the client can be reassured that the nature of the vast majority of their sequences will not be revealed.

We have developed an automated system for screening sequences in this way. The software, called BlackWatch, is consists of the custom, curated database of relevant sequences, a sophisticated web interface for submitting searches and viewing results, and core software that runs the BLAST sequence comparison software on the input sequences. The software interprets the search results, minimizing false positive matches, archiving the results and notifying staff by email if a match is found. The software has been in use at Blue Heron Biotechnology to screen sequences submitted for gene synthesis for more than a year. The details of the software are described in other Craic literature and an installation of the system can be accessed at this URL http://biotech.craic.com/blackwatch

Challenges

Sequence screening using BlackWatch offers one way for service companies to identify the presence sequences of hazardous agents. Our experience with the approach thus far is encouraging but there remain a number of challenges. Because of sequence conservation between organisms, particularly at the protein level, we do see false positive results. This is especially true with sequences from prokaryotes. For example, we have the entire genome of Yersinia pestis in the database. The strong conservation among ribosomal proteins between species means that a sequence for one of these from a harmless species will likely produce a match in a BlackWatch search. One approach to reducing false positives would be improve the curation of the database so as to only focus on genes that are involved in virulence or that encode toxins. Another would be to use a second database of sequences from harmless organisms and to identify false positives as matches to both.

Likewise we would like to minimize the number of false negatives. All sequences in our database are derived from those publicly available in GenBank and so we are inherently limited by that dataset. Some organisms, such as Yersinia, have their entire genome available. Others have relatively few genes characterized and in some cases, most notably the Anthrax genome, the sequences have not been released to GenBank. Clearly we can not find a match to a sequence that we do not have in our database.

False positive matches are likely to pose a significant problem in some applications of this technology. Each positive match will require a response from the service company in the form of contacting the client, stopping the synthesis, etc. If the false positive rate is too high then a significant effort will be expended dealing with spurious matches. This will lead to frustration, wasted effort and possibly lost business.

The nature of that response to a positive match poses perhaps the greatest challenge and one that we do not have a single clear answer to. If we find that a client has submitted a gene for synthesis from a hazardous agent then what should the service company do with that information ? Should they contact the CDC, or the FBI ? Should they contact the client for clarification about the intended use of the gene ? The existing notification mechanism for select agents involves the CDC but this is specific to the possession and transfer of agents as set out in 43 CFR part 73 and does not address the issues we have discussed here.

This question of 'What do we do if we find a hit?' is probably the major concern amongst the service companies that we have met with. The absence of a formal notification procedure places the onus on them to assess the implications of a positive database match and saddles them with potentially huge liability if they choose not inform the authorities of a possible threat. The CDC would seem to be the appropriate agency to notify but they are currently not set up to receive or evaluate this information.

Dialog between Industry and Government

We believe the issues of hazardous agents pose a serious problem for the biotechnology industry. A great deal of CDC and NIH attention is currently devoted to hazardous agents in terms of possession and transfer of the native organisms. This attention should be expanded to include the special issues raised here as they apply to genetic material and sequences derived from these organisms. In our experience the service companies are eager to engage the relevant agencies in dialog to better understand the issues. We wish to encourage this dialog and hope that we can play an important role in that process.

Deployment of BlackWatch Software

The approach we have implemented with BlackWatch is a practical solution that can be used today to address some of the risks discussed here. The software has been developed by Craic Computing with no external funding. One option for us is to market the software commercially, but this is a relatively small market that may not support the project at the level it demands. An alternative that we would like to pursue is to support the work through a contract from a government agency, such as CDC, NIAID or DARPA, and to provide the software at no or minimal cost to service companies. This approach would allow for the maximum deployment of the technology which in turn would provide the maximum level of protection against these risks within the industry.

This document was prepared by Dr. Robert Jones. For more information on this topic and the BlackWatch software please contact him at jones@craic.com

Craic Computing LLC, 911 East Pike Street, Suite 208, Seattle, WA 98122

URL	http://www.craic.com https://biotech.craic.com/blackwatch	Phone	206-860-7118
Email	blackwatch@craic.com	Fax	206-860-7119